AD-786 756

THE ECONOMICS OF DECISION MAKING

Thomas Richard Rice

Stanford University

Prepared for:

Office of Naval Research
Advanced Research Projects Agency
National Science Foundation

31 August 1974

AD786756

Research Report No. ELS-DA-74-2
May 1, 1973 to August 31, 1974

THE ECONOMICS OF DECISION MAKING

THOMAS RICHARD RICE

DECISION ANALYSIS PROGRAM

Professor Ronald A. Howard
Principal Investigator

DEPARTMENT OF ENGINEERING-ECONOMIC SYSTEMS

Stanford University
Stanford, California 94305

SPONSORSHIPS

Advanced Research Projects Agency, Human Resources Research Office, ARPA Order No. 2449, monitored by Engineering Psychology Programs, Office of Naval Research, Contract No.N00014-67-A-0112-0077(NR197-024)

National Science Foundation, NSF Grant GK-36491

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER EES-DA-74-2	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER AD 786 756
4. TITLE (and Subtitle) "The Economics of Decision Making"		5. TYPE OF REPORT & PERIOD COVERED Technical 5/1/73 to 8/31/74
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Thomas Richard Rice		8. CONTRACT OR GRANT NUMBER(s) N00014-67-A-0112-0077
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Board of Trustees of the Leland Stanford Junior University, c/o Office of Research Administrator, Encina Hall, Stanford, California 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 000000 ARPA Order #2449
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency Human Resources Research Office 1400 Wilson Blvd., Arlington, Virginia 22209		12. REPORT DATE August 31, 1974
		13. NUMBER OF PAGES pages 21
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) Engineering Psychology Programs, Code 455 Office of Naval Research 800 N. Quincy Street, Arlington, VA 22217		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

A paper submitted to MANAGEMENT SCIENCE/THEORY.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

DECISION ANALYSIS	DECISION THEORY	APPROXIMATE ANALYSES
DECISION-MAKING	VALUE OF INFORMATION	QUADRATIC PROBLEM

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The design of a decision analysis is itself a complex decision problem. In theory, each aspect of analysis, encoding the probability density functions of state variables, encoding the von Neuman-Morgenstern utility function, and computing profit lotteries is an experiment. The results of the experiments, the data, are used to update the probabilities in the primary decision problem. The economic value of the experiment is the well known value of imperfect information. (continued on reverse)

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

The drawback to the theoretical approach is that the data are functions. Practical methods for encoding prior distributions over functions do not exist. Therefore, the traditional approach is to parameterize the data.

Our approach is unique because we show that for an interesting class of decision problems, arbitrary parameterization is not necessary. The value of any data depends probabilistically only on the prior covariances of the posterior means. For independent state variables this quantity reduces to an estimate of how much the mean of a probability density function will shift during an experiment.

Thomas Richard Rice
Stanford University, 1974

ABSTRACT

The design of a decision analysis is itself a complex decision problem. In theory, each aspect of analysis, encoding the probability density functions of state variables, encoding the von Neuman-Morgenstern utility function, and computing profit lotteries is an experiment. The results of the experiments, the data, are used to update the probabilities in the primary decision problem. The economic value of the experiment is the well-known value of imperfect information.

The drawback to the theoretical approach is that the data are functions. Practical methods for encoding prior distributions over functions do not exist. Therefore, the traditional approach is to parameterize the data.

Our approach is unique because we show that for an interesting class of decision problems, arbitrary parameterization is not necessary. The value of any data depends probabilistically only on the prior covariances of the posterior means. For independent state variables this quantity reduces to an estimate of how much the mean of a probability density function will shift during an experiment.

THE VALUE OF DATA FOR A QUADRATIC DECISION PROBLEM

Approximate value of information calculations are essential in performing large decision analyses. Based on a preliminary analysis, the analyst must decide how to allocate his resources. His options include traditional experiments, such as market surveys and pilot plants, as well as additional analysis. The encoding of probability distributions and risk preference functions along with decision trees and Monte Carlo simulation are included in analysis.

All of these activities are directed towards updating our estimates of the outcome or value. Their worth depends on the prior assessment of how much they might change our decision and the subsequent gain in expected value. However, for complex decision problems finding the exact value of data requires an excessive amount of encoding and computation. The analyst needs to be able to make rapid, approximate value of information calculations.

This paper addresses the class of decision problems where the value function is approximately quadratic in both decision and state variables. The main result is that the value of an experiment depends only on the prior variance of the posterior mean. This is a tremendous simplification over the general case where the value of information depends on the prior probability distribution of the posterior probability distribution.

After we have proved the main result, we examine special cases to show that this is not a new idea but a generalization of an old one. The special cases and the discussion also identify the data required to operationalize the theorem.

1. Preliminaries

In this section we introduce inferential notation and the general terminology required to describe the decision problem.

Notation

Inferential notation explicitly conditions all probabilities on a state of information. The probability density function of a random variable x conditioned on the state of information $\mathcal{S}$ is denoted by

$$\{x|\mathcal{S}\} \quad (1.1)$$

We use $\int_x$ as a generalized summation operator; thus the k^{th} moment of x is

$$\langle x^k|\mathcal{S}\rangle = \int_x x^k\{x|\mathcal{S}\} \quad (1.2)$$

whether x is continuous or discrete. Inferential notation can be nested. For example,

$$\{\langle x|\mathcal{S}_2\rangle|\mathcal{S}_1\} \quad (1.3)$$

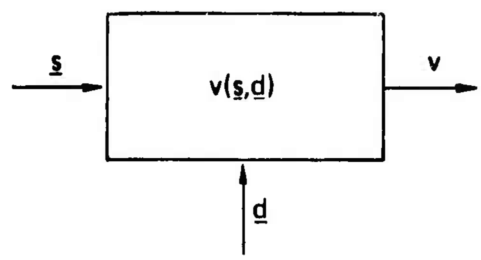
implies that the mean of $\{x|\mathcal{S}_2\}$ is a random variable given only $\mathcal{S}_1$.

In addition to inferential notation, we use the following matrix symbols:

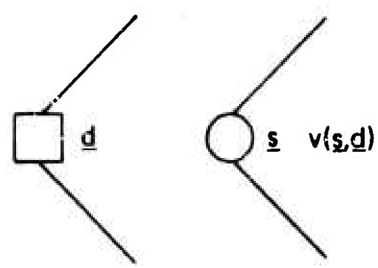
$\underline{a}$ or $a_i]$	The underscored lower case letter denotes a column vector with element a_i .
$\underline{A}$ or $[a_{ij}]$	The underscored capital letter denotes a square matrix with element a_{ij} .
$\underline{a}'$ or $\underline{A}'$	The prime denotes transposition .
$\langle\underline{a}|\mathcal{S}\rangle$ or $\langle a_i|\mathcal{S}\rangle]$	A probabilistic operation is applied to each component of a vector.

<u>The Basic Decision Problem</u>

The deterministic model illustrated in Fig. 1 relates the three elements of the basic decision problem. The decision variables $\underline{d}$ are set by the decision maker. The state variables $\underline{s}$ are set by nature. The value v is the output measure that we want to maximize. If both $\underline{s}$ and



(a) THE DETERMINISTIC MODEL



(b) THE PROBABILITY TREE

FIGURE 1 DESCRIPTION OF THE BASIC DECISION PROBLEM

$\underline{d}$ are known, we denote the decision that maximizes the value function $\underline{d}^+(\underline{s})$

$$(1.4) \qquad \underline{d}^+(\underline{s}) = \max_{\underline{d}}{}^{-1}\, v(\underline{s},\underline{d}) \ .$$

However, in the basic decision problem illustrated in Fig. 1(b), $\underline{d}$ must be set before $\underline{s}$ is observed. The possible outcomes are described by the probability density function $\{\underline{s}|\mathcal{E}\}$, where $\mathcal{E}$ is the state of information that represents the decision maker's prior knowledge and experience.

We assume that $\underline{s}$ independent of $\underline{d}$ in the sense that

$$(1.5) \qquad \{\underline{s}|\underline{d},\mathcal{E}\} = \{\underline{s}|\mathcal{E}\} \ .$$

This assumption is not restrictive. When the state variables are dependent on the decision variables the problem can normally be reformulated so that the dependence appears in the value function.

The basic decision problem under uncertainty is to maximize the expectation of v :

$$(1.6) \qquad \max_{\underline{d}} \int_{\underline{s}} v(\underline{s},\underline{d})\ \{\underline{s}|\mathcal{E}\}$$

The expansion rule from elementary probability theory is

$$(1.7) \qquad \langle x|\mathcal{E}\rangle = \int_{y} \langle x|y,\mathcal{E}\rangle\ \{y|\mathcal{E}\} \ .$$

Using this rule, we can show that the inferential symbol for the expectation in (1.6) is $\langle v|\underline{d},\mathcal{E}\rangle$:

$$(1.8) \qquad \langle v|\underline{d},\mathcal{E}\rangle = \int_{\underline{s}} \langle v|\underline{s},\underline{d},\mathcal{E}\rangle \{\underline{s}|\mathcal{E}\}$$

The expectations in (1.6) and (1.8) are the same since the expected value of v, given $\underline{s}$ and $\underline{d}$ is deterministically $v(\underline{s},\underline{d})$.

We define $\underline{d}^+(\mathcal{E})$ as the decision vector that maximizes the expected value of v:

$$(1.9) \qquad \underline{d}^+(\mathcal{E}) = \max_{\underline{d}}{}^{-1} \langle v|\underline{d},\mathcal{E}\rangle$$

If $\mathcal{S}$ represents some possible future state of information, we define $\underline{d}^*(\mathcal{S})$ as the intent to use $\underline{d}^+(\mathcal{S})$ when $\mathcal{S}$ becomes available.

<u>The Value of Information</u>

Suppose that an analysis or experiment will provide some data D. Then $(D,\mathcal{E})$ represents an improved state of information. We define the expected value of the data $\langle v_D|\mathcal{E}\rangle$:

$$(1.10) \qquad \langle v_D|\mathcal{E}\rangle = \langle v|\underline{d}^*(D,\mathcal{E}),\mathcal{E}\rangle - \langle v|\underline{d}^*(\mathcal{E}),\mathcal{E}\rangle$$

Since $\mathcal{E}$ is our prior information, $\underline{d}^+(\mathcal{E})$ is known and thus

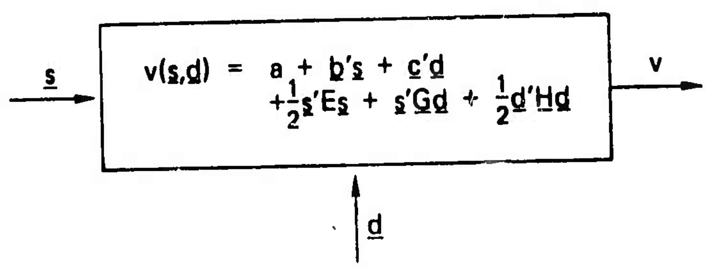
$$(1.11) \qquad \langle v|\underline{d}^*(\mathcal{E}),\mathcal{E}\rangle = \langle v|\underline{d}^+(\mathcal{E}),\mathcal{E}\rangle .$$

The first term in (1.10) is the key to the value of data. Given the data D we would find

$$(1.12) \qquad \underline{d}^+(D,\mathcal{E}) = \max_{\underline{d}}{}^{-1} \langle v|\underline{d},D,\mathcal{E}\rangle ,$$

which would result in the posterior expected value $\langle v|\underline{d}^+(D,\mathcal{E}),D,\mathcal{E}\rangle$. However, before D is revealed we must compute the prior expectation of this quantity:

$$(1.13) \qquad \langle v|\underline{d}^*(D,\mathcal{E}),D,\mathcal{E}\rangle = \langle\langle v|\underline{d}^+(D,\mathcal{E}),D,\mathcal{E}\rangle|\mathcal{E}\rangle$$



<u>NOTATION</u>

'	Transpose of a matrix
a	Constant
$\underline{b},\underline{c}$	Constant vectors
$\underline{s}$	State variable vector
$\underline{d}$	Decision variable vector
$\underline{E},\underline{G},\underline{H}$	Constant square matrices

FIGURE 2 THE QUADRATIC VALUE MODEL

2. The Value of Data for a Decision Problem with a Quadratic Value Function

In this section we find the expected value of data for the model illustrated in Fig. 2. The value function $v(\underline{s},\underline{d})$ is quadratic in the state vector $\underline{s}$ and the decision vector $\underline{d}$. The state variables have zero mean, and the decision variables are zero at the deterministic maximum :

$$(2.1) \qquad \langle \underline{s}|\mathcal{E}\rangle = \underline{0}$$

$$(2.2) \qquad \underline{d}_o^+ = \max_{\underline{d}}{}^{-1} v(\langle \underline{s}|\mathcal{E}\rangle,\underline{d}) = \underline{0}$$

These assumptions reduce algebraic complexity without sacrificing generality.

We write the quadratic value function as

$$(2.3) \qquad v(\underline{s},\underline{d}) = a + \underline{b}'\underline{s} + \underline{c}'\underline{d} + \frac{1}{2}\,\underline{s}'\underline{E}\,\underline{s} + \underline{s}'\underline{G}\,\underline{d} + \frac{1}{2}\,\underline{d}'\underline{H}\,\underline{d} .$$

The second-order necessary and sufficient conditions for $v(\underline{s},\underline{d})$ to have a maximum at $\langle \underline{s}|\mathcal{E}\rangle$ and $\underline{d}_o^+$ are that the gradient of v with respect to $\underline{d}$ $\nabla v(\langle \underline{s}|\mathcal{E}\rangle,\underline{d}_o^+)$ be zero and that the Hessian $\nabla^2 v(\langle \underline{s}|\mathcal{E}\rangle,\underline{d})$ be negative definite. Using (2.1) and (2.2) the gradient and Hessian at $\langle \underline{s}|\mathcal{E}\rangle$ and $\underline{d}_o^+$ are defined as :

$$(2.4) \qquad \nabla v(\langle \underline{s}|\mathcal{E}\rangle,\underline{d}_o^+) = \left\lfloor \frac{\partial v(\underline{0},\underline{0})}{\partial d_i} \right\rfloor$$

$$(2.5) \qquad \nabla^2 v(\langle \underline{s},\mathcal{E}\rangle,\underline{d}_o^+) = \left[\frac{\partial^2 v(\underline{0},\underline{0})}{\partial d_i \partial d_j} \right]$$

Applying (2.4) and (2.5) to the definition of $v(\underline{s},\underline{d})$ (2.3), we have :

$$\nabla v(\underline{0},\underline{0}) = \underline{c}' \tag{2.6}$$

$$\nabla^2 v(\underline{0},\underline{0}) = \underline{H} \tag{2.7}$$

Since the gradient in (2.6) must be the zero vector, our assumptions imply that $\underline{c}$ must also be the zero vector. From (2.7) we see that the Hessian does not vary with $\underline{s}$ and $\underline{d}$ for the quadratic. Therefore if the deterministic optimum $\underline{d}_o^+$ exists, $\underline{H}$ is negative definite and the value function has a global maximum with respect to $\underline{d}$ for any state vector $\underline{s}$.

Chronologically, we receive the data about the state variables. Then we set the decision vector, and finally nature sets the state variables. The state variables are independent of the decision variables but not necessarily independent of each other. We assume that the decision maker is risk-indifferent so that maximizing the value function is equivalent to maximizing the decision maker's von Neumann-Morgenstern utility function.

With these preliminatires we can state the theorem:

THEOREM: For the quadratic value function

$$v(\underline{s},\underline{d}) = a + \underline{b}'\underline{s} + \frac{1}{2}\,\underline{s}'\underline{E}\,\underline{s} + \underline{s}'\underline{G}\,\underline{d} + \frac{1}{2}\,\underline{d}'\underline{H}\,\underline{d} \tag{2.8}$$

where the Hessian $\underline{H}$ is negative definite, the value of any data D is

$$\langle v_D|\mathcal{E}\rangle = -\frac{1}{2}\langle\langle\underline{s}|D,\mathcal{E}\rangle'\,\underline{G}\,\underline{H}^{-1}\,\underline{G}'\,\langle\underline{s}|D,\mathcal{E}\rangle|\mathcal{E}\rangle\ . \tag{2.9}$$

PROOF: From (1.10) the value of the data is

$$\langle v_D|\mathcal{E}\rangle = \langle v|\underline{d}^*(D,\mathcal{E}),D,\mathcal{E}\rangle - \langle v|\underline{d}^*(\mathcal{E}),\mathcal{E}\rangle \,. \tag{2.10}$$

The proof is in two parts corresponding to the two terms of (2.10). First we determine the prior maximum expected value $\langle v|\underline{d}^*(\mathcal{E}),\mathcal{E}\rangle$; then we determine the expected value given the opportunity to maximize after the data is received $\langle v|\underline{d}^*(D,\mathcal{E}),D,\mathcal{E}\rangle$.

To find $\langle v|\underline{d}^*(\mathcal{E}),\mathcal{E}\rangle$ we start with the prior expected value $\langle v|\underline{d},\mathcal{E}\rangle$. Recalling that the expected value of the state variables are all zero, the prior expectation of (2.8) is

$$\langle v|\underline{d},\mathcal{E}\rangle = a + \frac{1}{2}\langle \underline{s}'\underline{E}\,\underline{s}|\mathcal{E}\rangle + \frac{1}{2}\,\underline{d}'\underline{H}\,\underline{d} \,. \tag{2.11}$$

The first-order necessary condition for $\langle v|\underline{d}^+(\mathcal{E}),\mathcal{E}\rangle$ to be an unconstrained maximum is that the gradient be zero at $d^+(\mathcal{E})$:

$$\nabla \langle v|\underline{d}^+(\mathcal{E}),\mathcal{E}\rangle = \underline{0}' \tag{2.12}$$

Taking the gradient of (2.11) and setting it to zero, we have

$$\underline{d}^{+\prime}(\mathcal{E})\,\underline{H} = \underline{0}'. \tag{2.13}$$

Since $\underline{H}$ is negative definite, $\underline{d}^+(\mathcal{E})$ must be the zero vector. Therefore (2.11) becomes

$$\langle v|\underline{d}^+(\mathcal{E}),\mathcal{E}\rangle = a + \frac{1}{2}\langle s'E\,s|\mathcal{E}\rangle \,. \tag{2.14}$$

Returning to the first term in (2.10), the expected value given data D is

$$\langle v|\underline{d},D,\mathcal{E}\rangle = a + \underline{b}'\langle\underline{s}|D,\mathcal{E}\rangle + \frac{1}{2}\langle\underline{s}'\underline{E}\ \underline{s}|D,\mathcal{E}\rangle$$

$$+ \langle\underline{s}|D,\mathcal{E}\rangle'\ \underline{G}\ \underline{d} + \frac{1}{2}\underline{d}'\underline{H}\ \underline{d}\ . \tag{2.15}$$

Maximizing (2.15) with respect to $\underline{d}$ we have

$$\nabla\langle v|\underline{d}^{+}(D,\mathcal{E}),D,\mathcal{E}\rangle = \langle\underline{s}|D,\mathcal{E}\rangle'\ \underline{G} + \underline{d}'\underline{H} = \underline{0}'\ . \tag{2.16}$$

Equation (2.16) implies that

$$\underline{d}^{+}(D,\mathcal{E}) = -\ \underline{H}^{-1}\ \underline{G}'\ \langle\underline{s}|D,\mathcal{E}\rangle\ . \tag{2.17}$$

Substituting (2.17) into (2.15), we have :

$$\langle v|\underline{d}^{+}(D,\mathcal{E}),D,\mathcal{E}\rangle = a + \underline{b}'\ \langle\underline{s}|D,\mathcal{E}\rangle + \frac{1}{2}\langle\underline{s}'\underline{E}\ \underline{s}|D,\mathcal{E}\rangle$$

$$-\ \langle\underline{s}|D,\mathcal{E}\rangle'\underline{G}\ \underline{H}^{-1}\underline{G}'\ \langle\underline{s}|D,\mathcal{E}\rangle + \frac{1}{2}\langle\underline{s}|D,\mathcal{E}\rangle'\ \underline{G}\ \underline{H}^{-1}\underline{G}'\ \langle\underline{s}|D,\mathcal{E}\rangle$$

$$= a + \underline{b}'\ \langle\underline{s}|D,\mathcal{E}\rangle + \frac{1}{2}\langle\underline{s}'\underline{E}\ \underline{s}|D,\mathcal{E}\rangle$$

$$+ \frac{1}{2}\langle\underline{s}|D,\mathcal{E}\rangle'\ \underline{G}\ \underline{H}^{-1}\underline{G}'\ \langle\underline{s}|D,\mathcal{E}\rangle \tag{2.18}$$

Recalling (1.13), the next step is to take the prior expectation of (2.18). We shall consider each term separately. Of course, expectation does not affect the value of the constant a . The prior expectation of the posterior mean is the prior mean :

$$\langle\langle\underline{s}|D,\mathcal{E}\rangle|\mathcal{E}\rangle = \langle\underline{s}|\mathcal{E}\rangle \tag{2.19}$$

Equation (2.19) is a direct application of the definition of conditional probability. Likewise, the third term becomes

$$\langle\langle\underline{s}'\underline{E}\ \underline{s}|D,\mathcal{E}\rangle|\mathcal{E}\rangle = \langle\underline{s}'\underline{E}\ \underline{s}|\mathcal{E}\rangle\ . \tag{2.20}$$

Applying these results to (2.18), we have

$$\langle v \lfloor \underline{d}^*(D,\mathcal{E}),\mathcal{E}\rangle = a + \frac{1}{2} \langle \underline{s}'\underline{E}\ \underline{s} | \mathcal{E}\rangle$$

$$(2.21) \qquad - \frac{1}{2} \left\langle \langle \underline{s} | D,\mathcal{E}\rangle' \ \underline{G}\ \underline{H}^{-1} \underline{G}' \ \langle \underline{s} | D,\mathcal{E}\rangle \middle| \mathcal{E} \right\rangle .$$

Finally, subtracting (2.14) from (2.21) the result is

$$(2.22) \qquad \langle v_D | \mathcal{E}\rangle = - \frac{1}{2} \left\langle \langle \underline{s} | D,\mathcal{E}\rangle' \ \underline{G}\ \underline{H}^{-1} \underline{G}' \ \langle \underline{s} | D,\mathcal{E}\rangle \middle| \mathcal{E} \right\rangle .$$

Q.E.D.

<u>Special Cases of the Theorem that Appear in the Literature</u>

Three special cases of the theorem (2.9) appear in the literature. Howard [2, p. 518] treats the case where $\underline{H}$ is diagonal and the data D is clairvoyance. DeGroot [1, p. 234] solves for $\underline{d}$, the estimate of the random variable $\underline{s}$ which minimizes a quadratic loss function. In our notation his problem is the case where

$$(2.23) \qquad \underline{E} = - \underline{H} ;$$

a , $\underline{b}$, and $\underline{G}$ are zero; and D is clairvouance. Raiffa and Schlaifer [4, p. 188] present the one-dimensional estimation problem without requiring the data to be clairvoyance.

3. <u>Discussion of the Value of Data for the Quadratic Problem</u>

An alternate expression for the theorem (2.9) is :

$$(3.1) \qquad \langle v_D | \mathcal{E}\rangle = - \frac{1}{2} \operatorname{trace} \underline{E}_{co} \underline{C}_D$$

where

$$(3.2) \qquad \underline{E}_{co} = \underline{G}\ \underline{H}^{-1}\ \underline{G}'$$

$$\underline{C}_D = \left[\langle\langle s_i | D, \mathcal{E}\rangle \langle s_j | D,\mathcal{E}\rangle | \mathcal{E} \rangle \right] \tag{3.3}$$

The trace of a matrix is the sum of its diagonal elements. The value of data has two major components. The basic decision problem is specified by $\underline{E}_{co}$, and the experiment is described by $\underline{C}_D$. We consider $\underline{E}_{co}$ and $\underline{C}_D$ briefly for the general case. Then for the case that is most common we discuss how $\underline{E}_{co}$ and $\underline{C}_D$ could be generated.

$\underline{E}_{co}$ follows directly from (3.2) for a true quadratic value function since the matrices $\underline{G}$ and $\underline{H}$ are specified. For a problem that is only approximately quadratic, $\underline{G}$ and $\underline{H}$ can be found by expanding $v(\underline{s},\underline{d})$ in a Taylor series about the point $(\langle \underline{s}|\mathcal{E}\rangle, \underline{d}^{+}(\mathcal{E})) = (\underline{0},\underline{0})$:

$$v(\underline{s},\underline{d}) = v(\underline{0},\underline{0}) + \frac{\partial v}{\partial s_i}\,\underline{s} + \frac{1}{2}\,\underline{s}'\left[\frac{\partial^2 v}{\partial s_i \partial s_j}\right]\underline{s}$$

$$+\ \underline{s}'\left[\frac{\partial^2 v}{\partial s_i \partial d_j}\right]\underline{d} + \frac{1}{2}\,\underline{d}'\left[\frac{\partial^2 v}{\partial d_i \partial d_j}\right]\underline{d} \tag{3.4}$$

The partial derivatives are all evaluated at the point $(\underline{0},\underline{0})$. Comparing (3.4) with (2.8), we see that $\underline{G}$ and $\underline{H}$ must be matrices of partial derivatives :

$$\underline{G} = \left[\frac{\partial^2 v}{\partial s_i \partial d_j}\right] \tag{3.5}$$

$$\underline{H} = \left[\frac{\partial^2 v}{\partial d_i \partial d_j}\right] \tag{3.6}$$

The partial derivatives at the operating point $(\underline{0},\underline{0})$ can be approximated from open loop sensitivities. One joint sensitivity is required for each possible pair of state and decision variables and for each possible pair of decision variables.

The elements of the matrix $\underline{C}_D$ are the expected product of the posterior means. Since the prior expectation of the posterior mean is zero, the elements are the covariances of the posterior means.

When the data is clairvoyance on the state variables $\underline{s}$, (3.1) reduces to

$$\langle v_c | \mathcal{E} \rangle = -\frac{1}{2} \text{ trace } \underline{E}_{co} \, \underline{C} \, . \tag{3.7}$$

If we consider the posterior means $\langle \underline{s} | D, \mathcal{E} \rangle$ as random variables, comparison of (3.7) and (3.1) implies that the value of data is the value of clairvoyance on the posterior means. In most practical problems the value of clairvoyance on the posterior mean is much easier to compute than the value of clairvoyance on the data itself.

<u>An Interesting Special Case</u>

The most interesting special case occurs when either $\underline{E}_{co}$ or $\underline{C}_D$ is a diagonal matrix. Then the value of data becomes

$$\langle v_D | \mathcal{E} \rangle = \sum_i \underline{g}_i' \, \underline{H}^{-1} \, \underline{g}_i \, {}^{v}\langle \langle s_i | D, \mathcal{E} \rangle | \mathcal{E} \rangle \tag{3.8}$$

where the vector $\underline{g}_i'$ is the i^{th} row of $\underline{G}$:

$$\underline{G} = \underline{g}_i'] \tag{3.9}$$

If the state variables are independent (3.8) is exactly equal to (3.1). Sufficient conditions for (3.7) to be a good approximation to (3.1) are that the diagonal elements dominate the off-diagonal elements of $\underline{E}_{co}$; that is, for each i and j :

$$\rho_{ij}^2 \ll \frac{(\underline{g}_i' \, \underline{H}^{-1} \, \underline{g}_i') \, (\underline{g}_j' \, \underline{H}^{-1} \, \underline{g}_j')}{(\underline{g}_i \, \underline{H}^{-1} \, \underline{g}_j)^2} \tag{3.10}$$

where ρ_{ij} is the correlation coefficient

$$(3.11)\quad \rho_{ij} = \langle\langle s_i|D,\mathcal{E}\rangle \langle s_j|D,\mathcal{E}\rangle|\mathcal{E}\rangle \Big/ \sqrt{\left({}^{v}\langle\langle s_i|D,\mathcal{E}\rangle|\mathcal{E}\rangle \, {}^{v}\langle\langle s_j|D,\mathcal{E}\rangle|\mathcal{E}\rangle\right)^{1/2}} \, .$$

Given $\underline{G}$, $\underline{H}$, and $\underline{C}_D$, these expressions tell us when the diagonal assumption holds. A more interesting question is whether we can avoid generating the entire matrices $\underline{G}$, $\underline{H}$ and $\underline{C}_D$. The answer is yes as shown below.

<u>Description of the Primary Problem Using Closed Loop Sensitivities</u>

We now show that the term $\underline{g}_i' \, \underline{H}^{-1} \, \underline{g}_i$ is the second partial derivative of compensation with respect to the i^{th} state variable :

$$(3.12)\qquad \underline{g}_i' \, \underline{H}^{-1} \, \underline{g}_i = \frac{\partial^2 v_{co}(s_i)}{\partial s_i^2}$$

where

$$(3.13)\qquad v_{co}(s_i) = v_c(s_i) - v_o(s_i)$$

The open loop sensitivity is evaluated by varying s_i while the other state variables and the decision variables remain constant. We denote the open loop sensitivity as

$$(3.14)\qquad v_o(s_i) = v(0,0,\ldots,s_i,\ldots,0,d_o^+) \, .$$

In closed loop sensitivity the state variables other than s_i remain fixed, but the decision is reoptimized for each s_i :

$$(3.15)\qquad v_c(s_i) = v(0,0,\ldots,s_i,\ldots,0,d^+(0,0,\ldots,s_i,\ldots,0))$$

To show that expression (3.12) is valid we evaluate $v_o(s_i)$ and

$v_c(s_i)$ for the quadratic value function (2.8) :

$$v_o(s_i) = a + b_i s_i + \frac{1}{2} e_{ii} s_i^2 \tag{3.16}$$

$$v_c(s_i) = \max_{\underline{d}}(a + b_i s_i + \frac{1}{2} e_{ii} s_i^2 + s_i \underline{g}_i' \underline{d} + \frac{1}{2} \underline{d} \underline{H} \underline{d})$$

$$= a + b_i s_i + \frac{1}{2} e_{ii} s_i^2 - \frac{1}{2} \underline{g}_i' \underline{H}^{-1} \underline{g}_i s_i^2 \tag{3.17}$$

Subtracting (3.16) from (3.17) the compensation is

$$v_{co}(s_i) = - \frac{1}{2} \underline{g}_i' \underline{H}^{-1} \underline{g}_i s_i^2 . \tag{3.18}$$

Therefore, by evaluating the compensation curves for the state variables, the need to find the matrices of partial derivatives $\underline{G}$ and $\underline{H}$ is eliminated.

The Description of the Data Generating Process through Preposterior Moments

The second component of (3.8) is $\overset{v}{<}\langle s_i | D, \mathcal{E}\rangle | \mathcal{E} >$ the prior variance of the posterior mean. To evaluate this term we use the theorem:

$$\overset{v}{<} s | \mathcal{E}> = \overset{v}{<}\langle s | D, \mathcal{E}\rangle | \mathcal{E}> + < \overset{v}{<} s | D, \mathcal{E}> | \mathcal{E}> \tag{3.19}$$

A proof of this theorem is given in Raiffa and Schlaifer [4, p. 106]. The theorem states that the prior variance $\overset{v}{<} s | \mathcal{E}>$ has two sources. The expected posterior variance $< \overset{v}{<} s | D, \mathcal{E}> | \mathcal{E}>$ is a residual variance which will not be resolved by the experiment that generates the data D . The prior variance of the posterior mean $\overset{v}{<}\langle s | D, \mathcal{E}\rangle | \mathcal{E}>$ is the portion of the prior variance that will be resolved by the experiment.

Sample Data

Expression (3.19) is best known for the case where data are N random samples from $\{s|\mathcal{E}\}$. First we consider the limiting cases of no samples and of infinite samples. Then we consider a finite number of samples.

When the data is the null experiment, $N = 0$, the prior and posterior states of information coincide. Therefore, we have

$$\langle {}^{v}\langle s|D,\mathcal{E}\rangle|\mathcal{E}\rangle = \langle {}^{v}\langle s|\mathcal{E}\rangle|\mathcal{E}\rangle = {}^{v}\langle s|\mathcal{E}\rangle \tag{3.20}$$

$$ {}^{v}\langle\langle s|D,\mathcal{E}\rangle|\mathcal{E}\rangle = {}^{v}\langle\langle s|\mathcal{E}\rangle|\mathcal{E}\rangle = 0 \ . \tag{3.21}$$

When the number of samples approaches infinity, the data is clairvoyance about s . The posterior probability density function will have all of its mass at a single point. Consequently, the preposterior moments are

$$\langle {}^{v}\langle s|D,\mathcal{E}\rangle|\mathcal{E}\rangle = \langle 0|\mathcal{E}\rangle = 0 \tag{3.22}$$

$$ {}^{v}\langle\langle s|D,\mathcal{E}\rangle|\mathcal{E}\rangle = {}^{v}\langle s|\mathcal{E}\rangle = {}^{v}\langle s|\mathcal{E}\rangle \tag{3.23}$$

To discuss (3.19) for finite N it is convenient to define the ratio r

$$r = {}^{v}\langle\langle s|D,\mathcal{E}\rangle|\mathcal{E}\rangle / {}^{v}\langle s|\mathcal{E}\rangle \ . \tag{3.24}$$

The limiting cases are $r = 0$ for the null experiment and $r = 1$ for clairvoyance.

A Bayesian must assign both r and $\{s|\mathcal{E}\}$ before he can calculate the expected value of sample information. For example, Raiffa and Schlaifer [4, p. 110] suggest assigning an equivalent sample size N' to the term ${}^{v}\langle\langle s|D,\mathcal{E}\rangle|\mathcal{E}\rangle$. Then for certain conditions the parameter r is

$$r = \frac{N}{N' + N} \quad (3.25)$$

Assigning other r or N' weights the prior information relative to the sample information.

<u>Experiments That Do Not Involve Sampling</u>

Encoding and modeling are analogous to sampling because they partially resolve uncertainty about the state variable s. Encoding the parameter r or equivalently ${}^{v}\langle\langle s|D,\mathcal{E}\rangle|\mathcal{E}\rangle$ should be no more difficult for these cases than for sampling.

4. Conclusion

Application of the approximate value of information is a two-step process. First, the value function must be approximated by a Taylor series. This step is routine given deterministic sensitivity data. The second step is to encode the prior covariances of the posterior means of the state variables. Practical applications are the result of careful modeling so that the encoding problem is tractable.

BIBLIOGRAPHY

[1] De Groot, Morris. Optimal Statistical Decisions, McGraw-Hill, 1970.

[2] Howard, Ronald A. "The Foundations of Decision Analysis," IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-4, No. 3, Sept. 1968.

[3] Howard, Ronald A. "Proximal Decision Analysis," Management Science, Vol. 17, No. 9, May 1971.

[4] Raiffa, Howard A., and Schlaifer, Robert. Applied Statistical Decision Theory, M.I.T. Press, Cambridge, Mass., 2nd Ed., 1968.